

# Separate Random Number Generation from Correlated Sources

Shengtian Yang\*

## Abstract

This work studies the problem of separate random number generation from correlated general sources with side information at the tester under the criterion of statistical distance. Tight one-shot lower and upper performance bounds are obtained using the random-bin approach. A refined analysis is further performed for two important random-bin maps. One is the pure-random-bin map that is uniformly distributed over the set of all maps (with the same domain and codomain). The other is the equal-random-bin map that is uniformly distributed over the set of all surjective maps that induce an equal or quasi-equal partition of the domain. Both of them are proved to have a doubly-exponential concentration of the performance of their sample maps. As an application, an open and transparent lottery scheme, using a random number generator on a public data source, is proposed to solve the social problem of scarce resource allocation. The core of the proposed framework of lottery algorithms is a permutation, a good rateless randomness extractor, whose existence is confirmed by the theoretical performance of equal-random-bin maps. This extractor, together with other important details of the scheme, ensures that the lottery scheme is immune to all kinds of fraud under some reasonable assumptions.

**Index Terms:** Correlated sources, information spectrum, lottery, randomness extractor, random number generation, random bins, side information, universal hashing.

## 1 Introduction

The problem of random number generation, namely extracting randomness from a random source, may date back to von Neumann [1], who designed a simple algorithm for simulating a fair coin by using a biased coin with unknown probability. So far, there has been a large body of research in this area, from theory to practice, but basically, all research can be classified into two categories.

The first category takes a model-based approach, assuming that the probability distribution of source is known or belongs to a family of probability distributions, e.g., [2–7] and the references therein. The main issues in this category are the estimation of model parameters and the

---

Version: 1.3.0-8ba126 (no. 201604282300).

This file was generated on May 2, 2016 by pdfTeX 1.40.12 with format LaTeX2e 2011/06/27.

Paper options: a4paper, onecolumn, 12pt, twoside.

geometry options: hscale = 0.84, vscale = 0.84, includehead, centering.

hyperref options: pagebackref, pdfstartview = FitH, hidelinks.

\*S. Yang is with the School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China (e-mail: [yangst@codlab.net](mailto:yangst@codlab.net)).

extraction of randomness from a source with known statistics. Traditional methods of lossless data compression with necessary modifications work well in this case.

In contrast, the second category of research, which is popular in computer science, does not make any assumption on the source, except that the source must have enough randomness to be extracted, e.g., [8, 9] and the references therein. A so-called min-entropy quantity, usually much less than the entropy of a source, is used to measure the amount of randomness of a source. The key issue in this category is designing a seeded (randomness) extractor that can extract a specified number of random bits from any data having enough randomness in terms of min-entropy. Its underlying theory is extensively related to universal hashing [10], or similarly, random bins, a technique widely used in distributed lossless source coding (i.e., Slepian-Wolf coding or separate coding of correlated sources) [11].

By the classification above, we show a clear duality between random number generation and lossless source coding. This duality is however not perfect in that the performance bound of extractors is characterized by min-entropy [12] whereas the achievable rate region of Slepian-Wolf coding is characterized by entropy [13]. Intuitively, since these two areas use the same principle of random bins, one may conjecture that there is a perfect duality between them. Does such a duality really exist? The answer is positive and is in fact partly known in recent research of random number generation from correlated sources (see Remarks 3.9 and 3.11), though not explicitly stated. In this paper, we will investigate this problem in depth. The contribution of this paper can be summarized as follows:

1) One-shot lower and upper bounds for the performance of separate random number generation from correlated sources with side information at the tester are derived. They are tight enough for the first- and second-order performance analysis. The upper bound even holds true for all 1-random-bin maps, a generalization of the classical random binning.

2) A phenomenon of doubly-exponential concentration of performance is proved for two kinds of random-bin maps, one of which is the equal-random-bin map, an important ensemble that has not received much attention in information theory. This result deepens our understanding of random number generation and gives us some useful hints on the design of random number generator.

3) A lottery scheme, using a random number generator on a public data source, is proposed to solve the social problem of scarce resource allocation. The core of the proposed framework of lottery algorithm is a permutation, a good rateless randomness extractor, whose existence is confirmed by the theoretical performance of equal-random-bin maps.

The rest of the paper is organized as follows: Section 2 formulates the problem of separate random number generation from correlated sources with side information at the tester. One-shot performance bounds as well as the achievable rate region of this problem is presented in Section 3. A refined analysis is performed in Section 4 for two important extractor ensembles, which turn out to have a sharp concentration of the performance of their individual extractors. In Section 5, we propose an open and transparent lottery scheme for resource allocation.

We close this section with some notations used throughout the paper. The ring of integers, the field of real numbers, and the finite field of order  $q$  are denoted by  $\mathbf{Z}$ ,  $\mathbf{R}$ , and  $\mathbf{F}_q$ , respectively. A subset of  $\mathbf{Z}$  (and similarly for  $\mathbf{R}$ ) is usually denoted by  $\mathbf{Z}_A := \mathbf{Z} \cap A$  for some set  $A$ . Then the set of integers from 1 to  $n$  is denoted  $\mathbf{Z}_{[1,n]}$ , and sometimes, if  $A = [0, +\infty)$  for example, we simply write  $\mathbf{Z}_{\geq 0}$  in place of  $\mathbf{Z}_{[0,+\infty)}$ . Since  $\mathbf{Z}_{[m,n]}$  will be frequently used, we further define its shorthand  $\llbracket m, n \rrbracket$ , where  $m$  and  $n$  may not be integers.

A sequence (or more generally, an indexed family) is a map  $x$  from an index set  $I$  into a value collection  $\mathcal{X}$ . Usually, we denote it by  $x = (x_i)_{i \in I}$  (or  $(x_i)_{i=1}^\infty$  if  $I = \mathbf{Z}_{>0}$ ), where  $x_i$  is the shorthand

of  $x(i)$ . The index  $i$  can also be written as the superscript of  $x$ , but we parenthesize it so that it is not confused with the power operation. When, for example,  $I = \mathbf{Z}_{>0}^3$ , we may further use the notation such as  $x = (x_{i,j}^{(k)})_{i,j,k \in \mathbf{Z}_{>0}}$ . A partial sequence of  $x$  is simply the restriction  $x|_A$  of map  $x$  to some  $A \subseteq I$ , and we write  $x_A = (x_i)_{i \in A}$  as shorthand, so that  $x$  becomes the further shorthand of  $x_I$ . Then, for example, for  $x = (x_i^{(j)})_{i,j \in \mathbf{Z}_{>0}}$ , the notations  $x_{\mathbf{Z}_{>0}}$ ,  $x^{(\mathbf{Z}_{>0})}$ ,  $x_{\mathbf{Z}_{>0}}^{(\mathbf{Z}_{>0})}$  all refer to the same sequence  $x$ , and  $x^{(j)}$  or  $x_{\mathbf{Z}_{>0}}^{(j)}$  refers to the partial sequence  $(x_i^{(j)})_{i=1}^\infty$ . For convenience of notation, we also define  $\Sigma(x) := \sum_{i \in I} x_i$  and  $\Pi(x) := \prod_{i \in I} x_i$  whenever these operations make sense for  $x$ . Then for a family  $\mathcal{X} = \mathcal{X}_I$  of alphabets, we write  $\Pi(\mathcal{X})$  in place of the cartesian product  $\prod_{i \in I} \mathcal{X}_i$ .

When performing probabilistic analysis, all objects of study are related to a basic probability space  $(\Omega, \mathfrak{A}, P)$  with  $\mathfrak{A}$  a  $\sigma$ -algebra in  $\Omega$  and  $P$  a probability measure on  $(\Omega, \mathfrak{A})$ . A random element in a measurable space  $(\Omega', \mathfrak{B})$  is a measurable mapping from  $\Omega$  into  $\Omega'$ . A random element uniformly distributed over  $\Omega'$  is denoted by  $U_{\Omega'}$ . For random elements  $X$  and  $Y$  in  $\Omega'$ , the *statistical distance* between  $P_X$  and  $P_Y$  is

$$d(P_X, P_Y) := \sup_{B \in \mathfrak{B}} |P_X(B) - P_Y(B)|.$$

When  $\Omega'$  is at most countable and  $\mathfrak{B}$  is the power set of  $\Omega'$ , we have

$$d(P_X, P_Y) = \frac{1}{2} \sum_{\omega' \in \Omega'} |P_X(\omega') - P_Y(\omega')|.$$

If  $d(P_X, P_Y) = 0$ , then  $X$  and  $Y$  have the same probability distribution and we write  $X \stackrel{d}{=} Y$ . A mixture distribution of  $P_X$  and  $P_Y$ , usually denoted  $\lambda P_X + (1 - \lambda)P_Y$ , is a convex combination of  $P_X$  and  $P_Y$  with some nonnegative weights  $\lambda$  and  $1 - \lambda$ .

## 2 Problem Formulation

Suppose that there are  $m+1$  correlated sources, indexed from 0 to  $m$ . With each source of positive index, there is associated a separate (seedless) extractor. Roughly speaking, we hope that the output data of all extractors are not only close to uniform but also almost independent of the source of index zero. Under this condition, we wonder how much randomness we can extract from the correlated sources.

The formal definition is given as follows, using the information-spectrum approach [5]:

**Definition 2.1** (Correlated sources). *Let  $\mathcal{X} = (\mathcal{X}_i)_{i \in \llbracket 0, m \rrbracket}$  denote an  $(m+1)$ -tuple of alphabets, where  $\mathcal{X}_0$  is at most countable and all the other  $\mathcal{X}_i$ 's are finite. An  $(m+1)$ -tuple of correlated (general) sources is a sequence  $X = (X_{\llbracket 0, m \rrbracket}^{(n)})_{n=1}^\infty$  of random elements*

$$X_{\llbracket 0, m \rrbracket}^{(n)} = (X_{i,j}^{(n)})_{i \in \llbracket 0, m \rrbracket, j \in \llbracket 1, n \rrbracket},$$

where  $i$  is the source index,  $j$  is the time index, and each  $X_{i,j}^{(n)}$  is a random element in  $\mathcal{X}_i$ . When  $X$  is stationary and memoryless, it can simply be identified with the random element  $(X_0^{(1)}, X_1^{(1)}, \dots, X_m^{(1)})$  (or  $(X_{0,j}^{(n)}, X_{1,j}^{(n)}, \dots, X_{m,j}^{(n)})$  for any  $j \in \llbracket 1, n \rrbracket$ ) in  $\mathcal{X}_0 \times \mathcal{X}_1 \times \dots \times \mathcal{X}_m$ .

**Definition 2.2** (Extractors). Let  $\varphi^{(n)} = (\varphi_i^{(n)} : \mathcal{X}_i^n \rightarrow \mathcal{Y}_i^{(n)})_{i \in \llbracket 1, m \rrbracket}$  be an  $m$ -tuple of (seedless) extractors,<sup>1</sup> with  $\mathcal{Y}^{(n)} = (\mathcal{Y}_i^{(n)})_{i \in \llbracket 1, m \rrbracket}$  denoting the  $m$ -tuple of output alphabets. The rate of each  $\varphi_i^{(n)}$  is defined as

$$R(\varphi_i^{(n)}) := \frac{1}{n} \ln |\mathcal{Y}_i^{(n)}|.$$

In the sequel, most of our analysis will be performed on some ensemble of extractors endowed with some probability measure, viz., a random extractor. For convenience, we assume from now on that every random extractor is independent of all the other random objects in question.

**Definition 2.3** (Achievable rate region). An  $m$ -tuple  $R = (R_i)_{i \in \llbracket 1, m \rrbracket}$  of rates is said to be achievable if there exists an  $m$ -tuple  $\varphi^{(n)}$  of extractors such that

$$\liminf_{n \rightarrow \infty} R(\varphi_i^{(n)}) \geq R_i \quad \text{for all } i \in \llbracket 1, m \rrbracket$$

and

$$\lim_{n \rightarrow \infty} d(P_{X_0^{(n)}(\varphi_i^{(n)}(X_i^{(n)}))_{i \in \llbracket 1, m \rrbracket}}, P_{X_0^{(n)}} P_{U_{\Pi(\mathcal{Y}^{(n)})}}) = 0.$$

Then the achievable rate region of  $X$  is

$$\mathcal{R}(X) := \{R \in \mathbf{R}_{\geq 0}^m : R \text{ is achievable}\}.$$

### 3 One-Shot Bounds and Achievable Rate Region

In this section, we will derive one-shot lower and upper performance bounds as well as the achievable rate region  $\mathcal{R}(X)$ . Since random bins are our main tool for constructing extractors, we first give an exact definition of random bins, which is more general than needed.

**Definition 3.1.** Let  $F$  be a random map from  $A$  to  $B$ , with  $A$  and  $B$  both finite. The map  $F$  is said to be a  $(Q, \gamma)$ -random-bin map for some probability distribution  $Q$  on  $B$  and some  $\gamma \geq 0$  if for distinct  $x, y \in A$  and  $z \in B$ ,

$$P\{F(x) = z\} = Q(z)$$

and

$$P\{F(x) = z, F(y) = z\} \leq \gamma Q(z)^2. \tag{1}$$

When  $Q$  is the uniform distribution on  $B$ , we simply say that  $F$  is a  $\gamma$ -random-bin map.

Below we provide some examples of  $\gamma$ -random-bin maps, which will be needed later on.

**Example 3.2** (Pure random binning). A pure-random-bin map  $F$  is a random map uniformly distributed over the set of all maps of  $A$  into  $B$ . It is clear that  $F$  is a 1-random-bin map.

**Example 3.3** (Equal random binning). A map  $f$  of  $A$  onto  $B$  is said to be an equal-bin map if

$$\left\lfloor \frac{|A|}{|B|} \right\rfloor \leq |f^{-1}(z)| \leq \left\lceil \frac{|A|}{|B|} \right\rceil$$

---

<sup>1</sup> There are two kinds of randomness extractors, seeded and seedless, but there is no substantial difference between them, since a general source combined with a random seed is again a general source.

for all  $z \in B$ . An equal-random-bin map  $F$  is a random map uniformly distributed over the set of all equal-bin maps from  $A$  onto  $B$ . It is clear that  $F \stackrel{d}{=} U_{S_B} \circ f \circ U_{S_A}$ , where  $S_A$  denotes the symmetric group of all permutations of  $A$  and  $f$  is an arbitrary fixed equal-bin map from  $A$  onto  $B$ . It is easy to figure out that for distinct  $x, y \in A$  and  $z \in B$ ,

$$P\{F(x) = z\} = \frac{1}{|B|}$$

and

$$P\{F(x) = z, F(y) = z\} = \frac{(|A| - r)(|A| - |B| + r)}{|A|(|A| - 1)|B|^2} \leq \frac{1}{|B|^2},$$

where  $r = |A| \bmod |B|$ . Therefore,  $F$  is a 1-random-bin map.

**Example 3.4** (Affine random binning). Suppose that  $A = \mathbf{F}_q^m$  and  $B = \mathbf{F}_q^n$ . Let  $C$  be an  $(m, n, k)$  maximum-rank-distance (MRD) code over  $\mathbf{F}_q$  with  $1 \leq k \leq \min\{m, n\}$  [14, 15]. An affine-random-bin map  $A_C$  is a random affine map from  $\mathbf{F}_q^m$  to  $\mathbf{F}_q^n$  given by  $v \mapsto vU_C + U_{\mathbf{F}_q^n}$  (where both  $\mathbf{F}_q^m$  and  $\mathbf{F}_q^n$  are regarded as row-vector spaces). By [16, Theorem 2.5 and Proposition 5],  $A_C$  is a 1-random-bin map. In particular, the set  $\mathbf{F}_q^{m \times n}$  of all  $m \times n$  matrices over  $\mathbf{F}_q$  is an  $(m, n, \min\{m, n\})$  MRD code, so that  $A_{\mathbf{F}_q^{m \times n}}$  is a 1-random-bin map.

**Example 3.5** (Binary linear random binning). Suppose that  $A = \mathbf{F}_2^m$  and  $B = \mathbf{F}_2^n$ . Let  $C$  be an  $(m, n, k)$  MRD code over  $\mathbf{F}_2$  with  $2 \leq k \leq \min\{m, n\}$ . A binary linear-random-bin map  $L_C$  is a random linear map from  $\mathbf{F}_2^m$  to  $\mathbf{F}_2^n$  given by  $v \mapsto vU_C$ . By [16, Theorem 3.6],  $L_C|_{\{0\}^c}$  is a 1-random-bin map. In particular,  $L_{\mathbf{F}_2^{m \times n}}|_{\{0\}^c}$  is a 1-random-bin map when  $m, n \geq 2$ .

For  $(Q, \gamma)$ -random-bin maps, we have the following important property:

**Theorem 3.6.** Let  $\mathcal{X} = \mathcal{X}_{[0, m]}$  and  $\mathcal{Y} = \mathcal{Y}_{[1, m]}$  be two families of alphabets with  $\mathcal{X}_0$  at most countable and all the other alphabets finite. Let  $X = (X_i)_{i \in [0, m]}$  be a random element in  $\Pi(\mathcal{X})$  and  $Y = (Y_i)_{i \in [1, m]}$  a random element in  $\Pi(\mathcal{Y})$  with all  $Y_i$ 's mutually independent (and also independent of any other random object). Let  $F = (F_i : \mathcal{X}_i \rightarrow \mathcal{Y}_i)_{i \in [1, m]}$  be an  $m$ -tuple of random maps such that each  $F_i$  is an independent  $(P_{Y_i}, \gamma_i)$ -random-bin map, where  $\gamma = (\gamma_i)_{i \in [1, m]}$  is an  $m$ -tuple of real numbers. Then, for  $a \in \mathcal{X}_0$  and  $y \in \Pi(\mathcal{Y})$ ,

$$\mathbb{E}[\mu_X(F, a, y)] = P_{X_0}(a)P_Y(y)$$

and

$$\text{Var}(\mu_X(F, a, y)) \leq (\Pi(\gamma) - 1)P_{X_0}(a)^2P_Y(y)^2 + \sum_{\substack{S \subseteq [1, m] \\ S \neq \emptyset}} \Pi(\gamma_{S^c})P_Y(y)P_{Y_{S^c}}(y_{S^c})\nu_X(a, S),$$

where

$$\begin{aligned} \mu_X(F, a, y) &:= P[X_0 = a, (F_i(X_i) = y_i)_{i \in [1, m]} \mid F], \\ \nu_X(a, S) &:= \sum_{u \in \Pi(\mathcal{X}_S)} P_{X_0 X_S}(a, u)^2, \end{aligned}$$

and  $S^c = [1, m] \setminus S$ .<sup>2</sup>

---

<sup>2</sup> This convention will be used in the sequel for subsets of index set  $[1, m]$  of correlated sources.

*Proof.* Using the property of conditional expectation, we have

$$\begin{aligned} \mathbb{E}[\mu_X(F, a, y)] &= P\{X_0 = a, (F_i(X_i) = y_i)_{i \in \llbracket 1, m \rrbracket}\} \\ &= \mathbb{E}[P\{X_0 = a, (F_i(X_i) = y_i)_{i \in \llbracket 1, m \rrbracket} \mid X\}] \\ &= \mathbb{E}\left[1\{X_0 = a\} \prod_{i=1}^m P_{Y_i}(y_i)\right] = P_{X_0}(a)P_Y(y). \end{aligned}$$

Since  $\text{Var}(\mu_X(F, a, y)) = \mathbb{E}[\mu_X(F, a, y)^2] - (\mathbb{E}[\mu_X(F, a, y)])^2$ , it suffices to calculate the first term at the right-hand side. We have

$$\begin{aligned} &\mathbb{E}[\mu_X(F, a, y)^2] \\ &= \mathbb{E}\left[\left(\sum_{x \in \mathcal{X}} P_X(x) 1\{x_0 = a, (F_i(x_i) = y_i)_{i \in \llbracket 1, m \rrbracket}\}\right)^2\right] \\ &\leq \sum_{S \subseteq \llbracket 1, m \rrbracket} \sum_{u \in \Pi(\mathcal{X}_S)} \sum_{\substack{v, w \in \Pi(\mathcal{X}_{S^c}) \\ (v_i \neq w_i)_{i \in S^c}}} P_{X_0 X_S X_{S^c}}(a, u, v) P_{X_0 X_S X_{S^c}}(a, u, w) P_Y(y) \Pi(\gamma_{S^c}) P_{Y_{S^c}}(y_{S^c}) \quad (2) \\ &\leq \sum_{S \subseteq \llbracket 1, m \rrbracket} \Pi(\gamma_{S^c}) P_Y(y) P_{Y_{S^c}}(y_{S^c}) \sum_{u \in \Pi(\mathcal{X}_S)} \left(\sum_{v \in \Pi(\mathcal{X}_{S^c})} P_{X_0 X_S X_{S^c}}(a, u, v)\right)^2 \\ &= \sum_{S \subseteq \llbracket 1, m \rrbracket} \Pi(\gamma_{S^c}) P_Y(y) P_{Y_{S^c}}(y_{S^c}) \sum_{u \in \Pi(\mathcal{X}_S)} P_{X_0 X_S}(a, u)^2, \end{aligned}$$

where (2) follows from the expansion of the product and the inequality

$$\begin{aligned} &\mathbb{E}[1\{x_0 = a, F_S(u) = y_S, F_{S^c}(v) = y_{S^c}\} 1\{x_0 = a, F_S(u) = y_S, F_{S^c}(w) = y_{S^c}\}] \\ &= \mathbb{E}[1\{x_0 = a\} 1\{F_S(u) = y_S\} 1\{F_{S^c}(v) = y_{S^c}, F_{S^c}(w) = y_{S^c}\}] \\ &\leq 1\{x_0 = a\} P_{Y_S}(y_S) \Pi(\gamma_{S^c}) [P_{Y_{S^c}}(y_{S^c})]^2 = 1\{x_0 = a\} P_Y(y) \Pi(\gamma_{S^c}) P_{Y_{S^c}}(y_{S^c}), \end{aligned}$$

so that

$$\text{Var}(\mu_X(F, a, y)) \leq (\Pi(\gamma) - 1) P_{X_0}(a)^2 P_Y(y)^2 + \sum_{\substack{S \subseteq \llbracket 1, m \rrbracket \\ S \neq \emptyset}} \Pi(\gamma_{S^c}) P_Y(y) P_{Y_{S^c}}(y_{S^c}) \nu_X(a, S). \quad \square$$

A useful consequence follows immediately from Theorem 3.6.

**Corollary 3.7.** *Under the same condition as Theorem 3.6,*

$$\mathbb{E}[d(X, Y \mid F)] \leq \frac{1}{2} \sum_{a \in \mathcal{X}_0, y \in \mathcal{Y}} \left[ (\Pi(\gamma) - 1) P_{X_0}(a)^2 P_Y(y)^2 + \sum_{\substack{S \subseteq \llbracket 1, m \rrbracket \\ S \neq \emptyset}} \Pi(\gamma_{S^c}) P_Y(y) P_{Y_{S^c}}(y_{S^c}) \nu_X(a, S) \right]^{\frac{1}{2}}.$$

where  $d(X, Y \mid F) := d(P_{X_0(F_i(X_i))_{i \in \llbracket 1, m \rrbracket}}[\cdot \mid F], P_{X_0} P_Y)$  is a  $\sigma(F)$ -measurable random variable.

*Proof.* It is clear that

$$\begin{aligned}
\mathbb{E}[d(X, Y \mid F)] &= \mathbb{E} \left[ \frac{1}{2} \sum_{a \in \mathcal{X}_0, y \in \mathcal{Y}} |\mu_X(F, a, y) - P_{X_0}(a)P_Y(y)| \right] \\
&= \frac{1}{2} \sum_{a \in \mathcal{X}_0, y \in \mathcal{Y}} \mathbb{E}[|\mu_X(F, a, y) - P_{X_0}(a)P_Y(y)|] \\
&\leq \frac{1}{2} \sum_{a \in \mathcal{X}_0, y \in \mathcal{Y}} \sqrt{\text{Var}(\mu_X(F, a, y))},
\end{aligned} \tag{3}$$

which combined with Theorem 3.6 establishes the corollary, where (3) follows from Jensen's inequality.  $\square$

We are now in the position to derive tight one-shot upper and lower bounds, which are the crux of the proof of the achievable rate region (Theorem 3.12) and are useful in a refined analysis for special random-bin maps (Section 4) or in the second-order performance analysis.

**Lemma 3.8.** *Let  $X = X_{\llbracket 0, m \rrbracket}$  be an  $(m+1)$ -tuple of random elements in  $\Pi(\mathcal{X})$  with  $\mathcal{X} = \mathcal{X}_{\llbracket 0, m \rrbracket}$ . Let  $\Phi = \{\Phi_i : \mathcal{X}_i \rightarrow \mathcal{Y}_i\}_{i \in \llbracket 1, m \rrbracket}$  be an  $m$ -tuple of 1-random-bin extractors. Then for  $r > 1$ ,*

$$\mathbb{E}[d(X \mid \Phi)] \leq P\{T_X(X) \notin A_r(\mathcal{Y})\} + \frac{\sqrt{2^m - 1}}{2} r^{-1/2},$$

where

$$d(X \mid \Phi) := d(X, U_{\Pi(\mathcal{Y})} \mid \Phi), \tag{4}$$

$$T_X(x) := (T_X^{(S)}(x))_{\emptyset \neq S \subseteq \llbracket 1, m \rrbracket} = \left( \frac{1}{P_{X_S|X_0}(x_S \mid x_0)} \right)_{\emptyset \neq S \subseteq \llbracket 1, m \rrbracket}, \tag{5}$$

$$A_r(\mathcal{Y}) := \prod_{\emptyset \neq S \subseteq \llbracket 1, m \rrbracket} I_{r|\Pi(\mathcal{Y}_S)|}, \tag{6}$$

$$I_t := (t, +\infty). \tag{7}$$

*Proof.* Let  $\lambda = P(G)$  with  $G = \{\omega : T_X(X) \in A_r(\mathcal{Y})\}$ . Since the lemma holds trivially for  $\lambda = 0$ , we suppose that  $\lambda > 0$ . Then

$$P_X = \lambda P_V + (1 - \lambda) P_W,$$

where  $P_V = P_{X|1_G=1}$  and  $P_W = P_{X|1_G=0}$  with  $1_G$  denoting the indicator function of  $G$ . In particular,  $P_V(x) \leq P_X(x)/\lambda$  for all  $x \in \Pi(\mathcal{X})$ . It is clear that

$$\begin{aligned}
d(X \mid \Phi) &\leq \lambda d(V \mid \Phi) + (1 - \lambda) d(W \mid \Phi) \\
&\leq \lambda d(V \mid \Phi) + 1 - \lambda,
\end{aligned}$$

where the first inequality is due to Proposition A.1. From Corollary 3.7, it follows that

$$\begin{aligned}
\mathbb{E}[d(V \mid \Phi)] &\leq \frac{1}{2} \sum_{a \in \mathcal{X}_0, y \in \mathcal{Y}} \left( \sum_{\emptyset \neq S \subseteq \llbracket 1, m \rrbracket} \frac{\nu_V(a, S)}{|\Pi(\mathcal{Y})| |\Pi(\mathcal{Y}_{S^c})|} \right)^{\frac{1}{2}} \\
&= \frac{1}{2} \sum_{a \in \mathcal{X}_0} \left( \sum_{\emptyset \neq S \subseteq \llbracket 1, m \rrbracket} |\Pi(\mathcal{Y}_S)| \nu_V(a, S) \right)^{\frac{1}{2}}.
\end{aligned}$$



Since  $V \in X(G)$  almost surely, we have  $T_X^{(S)}(V) > r|\Pi(\mathcal{Y}_S)|$  almost surely, so that  $P_{X_S|X_0}(V_S | V_0) < p = 1/(r|\Pi(\mathcal{Y}_S)|)$  almost surely, and hence

$$\begin{aligned}
\nu_V(a, S) &= \sum_{u: P_{X_S|X_0}(u|a) < p} P_{V_0 V_S}(a, u)^2 \\
&\leq \sum_{u: P_{X_S|X_0}(u|a) < p} \left( \sum_{v \in \Pi(\mathcal{X}_{S^c})} \frac{P_{X_0 X_S X_{S^c}}(a, u, v)}{\lambda} \right)^2 \\
&= \lambda^{-2} \sum_{u: P_{X_S|X_0}(u|a) < p} P_{X_0 X_S}(a, u)^2 \\
&= \lambda^{-2} P_{X_0}(a)^2 \sum_{u: P_{X_S|X_0}(u|a) < p} P_{X_S|X_0}(u | a)^2 \\
&< p \lambda^{-2} P_{X_0}(a)^2 = r^{-1} |\Pi(\mathcal{Y}_S^{(n)})|^{-1} \lambda^{-2} P_{X_0}(a)^2.
\end{aligned}$$

Then

$$\mathbb{E}[d(V | \Phi)] \leq \frac{\sqrt{2^m - 1}}{2\lambda} r^{-1/2},$$

so that

$$\mathbb{E}[d(X | \Phi)] \leq \frac{\sqrt{2^m - 1}}{2} r^{-1/2} + 1 - \lambda.$$

The proof is complete.  $\square$

**Remark 3.9.** *The proof of Lemma 3.8 include three important ideas:*

- 1) Using 1-random-bin maps as randomness extractors (Definition 3.1 with  $Q = P_{U_B}$  and  $\gamma = 1$ ). 1-random-bin map is a generalization of the random-bin idea (i.e., the pure-random-bin map in Example 3.2) extensively used in information theory (e.g., [17, 18]) and is similar to the idea of universal hashing [10].

*Our definition is stronger than the universal<sub>2</sub> condition of [10], which, in probability form, only requires the collision probability*

$$P\{F(x) = F(y)\} \leq \frac{1}{|B|}$$

*for any distinct  $x, y \in A$ . However, our definition is weaker than another form of universal hashing [19, Definition 1.1], which requires*

$$P\{F(x) = z, F(y) = u\} = \frac{1}{|B|^2}$$

*for any distinct  $x, y \in A$  and any  $z, u \in B$ .*

*Thanks to this distinction (i.e., inequality (1)), 1-random-bin maps include some important ensembles that have not received much attention in the research of random number generation, such as equal-random-bin maps (Example 3.3) and binary linear-random-bin maps (Example 3.5).*



- 2) Using Jensen's inequality and the variance of output probability of 1-random-bin map to obtain an upper bound of statistical distance between the distribution of 1-random-bin extractor output and some fixed target distribution (Corollary 3.7). *Similar approaches have already appeared in previous research, e.g., [19, Lemma 4.4] (using Cauchy-Schwarz inequality), [17, Eq. (32)] (using Cauchy-Schwarz inequality and Jensen's inequality), [20, Appendix A] (using Cauchy-Schwarz inequality).*
- 3) Utilizing the convexity of statistical distance, i.e., Proposition A.1. *Simply using idea (2) does not work for a general source that has a subset of high-probability elements whose total probability is negligible. This difficulty can be overcome by partition the alphabet of source into two parts: one part including all small-probability elements that can effectively be handled by idea (2), and the other part including all remaining elements. The success of this approach is ensured by the convexity of statistical distance and the boundedness of statistical distance. Similar approaches have already been used in, e.g., [17, 20], both using the triangle inequality.*

In summary, the novelty of Lemma 3.8 is a systematic combination of the above ideas, especially idea (1), which enables a refined analysis of equal-random-bin maps in Sec. 4.

**Lemma 3.10.** *Let  $X = X_{[0,m]}$  be an  $(m+1)$ -tuple of random elements in  $\Pi(\mathcal{X})$  with  $\mathcal{X} = \mathcal{X}_{[0,m]}$ . Let  $\varphi = (\varphi_i : \mathcal{X}_i \rightarrow \mathcal{Y}_i)_{i \in [1,m]}$  be an  $m$ -tuple of extractors. Then, for  $0 < r < 1$ ,*

$$d(X \mid \varphi) \geq P\{T_X(X) \notin A_r(\mathcal{Y})\} - (2^m - 1)r,$$

where  $d(\cdot \mid \cdot)$ ,  $T_X$ , and  $A_r$  are defined by (4), (5), and (6), respectively.

*Proof.* Let  $G = \{\omega : T_X(X) \in A_r(\mathcal{Y})\}$  and  $B = (X_0(\varphi_i(X_i))_{i \in [1,m]})(G^c)$ . Then by definition,

$$\begin{aligned} d(X \mid \varphi) &\geq P_{X_0(\varphi_i(X_i))_{i \in [1,m]}}(B) - P_{X_0 U_{\Pi(\mathcal{Y})}}(B) \\ &\geq P(G^c) - P_{X_0 U_{\Pi(\mathcal{Y})}}(B). \end{aligned}$$

The proof will be finished if we find an upper bound of  $P_{X_0 U_{\Pi(\mathcal{Y})}}(B)$ . It is clear that

$$G = \bigcap_{\emptyset \neq S \subseteq [1,m]} C_S$$

with  $C_S = \{\omega : T_X^S(X) \in I_{r|\Pi(\mathcal{Y}_S)|}\}$ , so that

$$B = \bigcup_{\emptyset \neq S \subseteq [1,m]} (X_0(\varphi_i(X_i))_{i \in [1,m]})(C_S^c),$$

and hence

$$\begin{aligned} P_{X_0 U_{\Pi(\mathcal{Y})}}(B) &\leq \sum_{\emptyset \neq S \subseteq [1,m]} P_{X_0 U_{\Pi(\mathcal{Y})}}((X_0(\varphi_i(X_i))_{i \in [1,m]})(C_S^c)) \\ &\leq \sum_{\emptyset \neq S \subseteq [1,m]} P_{X_0 U_{\Pi(\mathcal{Y}_S)}}((X_0(\varphi_i(X_i))_{i \in S})(C_S^c)) \\ &= \sum_{\emptyset \neq S \subseteq [1,m]} \sum_{a \in \mathcal{X}_0} P_{X_0}(a) P_{U_{\Pi(\mathcal{Y}_S)}}(D_S(a)) \\ &= \sum_{\emptyset \neq S \subseteq [1,m]} \sum_{a \in \mathcal{X}_0} P_{X_0}(a) \frac{|D_S(a)|}{|\Pi(\mathcal{Y}_S)|} \\ &\leq (2^m - 1)r, \end{aligned}$$

where

$$D_S(a) = \{(\varphi_i(X_i(\omega)))_{i \in S} : \omega \in C_S^c, X_0(\omega) = a\}$$

and

$$|D_S(a)| \leq |\{X_S(\omega) : \omega \in C_S^c, X_0(\omega) = a\}| \leq r |\Pi(\mathcal{Y}_S)|$$

because  $1/P_{X_S|X_0}(X_S(\omega) | a) = T_X^{(S)}(X(\omega)) \leq r |\Pi(\mathcal{Y}_S)|$  for all  $\omega \notin C_S$  with  $X_0(\omega) = a$ .  $\square$

**Remark 3.11.** *The proof of Lemma 3.10 is mainly based on the idea of the proof of [5, Lemma 2.1.2], and is a generalization of [5, Lemma 2.1.2] and [21, Lemma 1] for uniform target distribution. With Lemma 3.10, we can even obtain an exact second-order performance, for example, in the memoryless case, which cannot be obtained by [5, Lemma 2.1.2] or [21, Lemma 1].*

With the one-shot bounds established above, we can easily obtain the achievable rate region  $\mathcal{R}(X)$ , whose proof is omitted here.

**Theorem 3.12.** *For an  $(m+1)$ -tuple  $X$  of correlated sources (see Definitions 2.1–2.3),*

$$\mathcal{R}(X) = \mathcal{R}_i(X) := \{R \in \mathbf{R}_{\geq 0}^m : \Sigma(R_S) \leq \underline{\mathcal{H}}(X_S | X_0) \text{ for all nonempty } S \subseteq \llbracket 1, m \rrbracket\},$$

where

$$\underline{\mathcal{H}}(X_S | X_0) := \text{p-lim inf}_{n \rightarrow \infty} \frac{1}{n} \ln \frac{1}{P_{X_S^{(n)}|X_0^{(n)}}(X_S^{(n)} | X_0^{(n)})}$$

is the spectral inf-entropy rate of  $X_S$  given  $X_0$ .<sup>3</sup>

Theorem 3.12 is very general. To understand its significance, we consider its consequences in some special cases.

**Corollary 3.13.** *For a pair  $X = (X_1^{(n)}, X_2^{(n)})_{n=1}^\infty$  of correlated stationary memoryless sources,*

$$\mathcal{R}(X) = \{(R_1, R_2) \in \mathbf{R}_{\geq 0}^2 : R_1 \leq H(X_1^{(1)}), R_2 \leq H(X_2^{(1)}), R_1 + R_2 \leq H(X_1^{(1)}, X_2^{(1)})\},$$

where  $H(X) := \mathbb{E}[-\ln P_X(X)]$  is the entropy of  $X$ .

Corollary 3.13 provides a perfect dual of the classical Slepian-Wolf coding theorem [13], as shown in Fig. 1. This result was first discovered by Oohama in the proof of his paper [17, 23]. But it was not explicitly stated, because the problem he studied only requires that the information at each source be separately encoded and then be jointly processed at the receiver for generating random numbers.

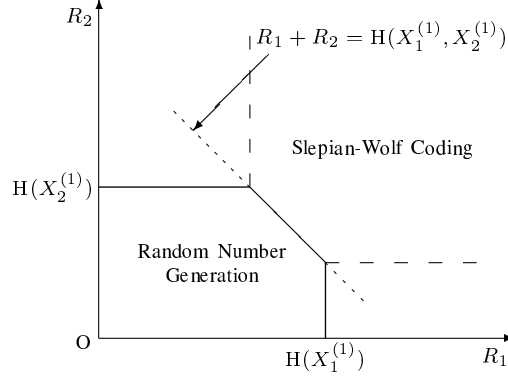
The direct part of Corollary 3.13 but with stationary memoryless side information was also proved recently in [18], whose method is also valid for the direct-part proof of correlated general sources. Thus, the method of [18, Theorem 1] and [21, Lemma 1] can also be used to establish Theorem 3.12, though, to the author's knowledge, Theorem 3.12 has never been formally proved. For the duality in this general case, the reader may compare Theorem 3.12 with [24, Theorem 1] or [5, Theorem 7.2.1].

**Corollary 3.14.** *For a pair  $X = (X_0^{(n)}, X_1^{(n)})_{n=1}^\infty$  of correlated sources,*

$$\mathcal{R}(X) = \{R_1 \in \mathbf{R}_{\geq 0} : R_1 \leq \underline{\mathcal{H}}(X_1 | X_0)\}.$$

---

<sup>3</sup> For a sequence  $(Z_n)_{n=1}^\infty$  of real-valued random variables  $Z_n$ ,  $\text{p-lim inf}_{n \rightarrow \infty} Z_n := \sup\{r : \lim_{n \rightarrow \infty} P\{Z_n < r\} = 0\}$  is the *limit inferior in probability* of  $(Z_n)_{n=1}^\infty$  [5, 22].



**Figure 1:** Achievable rate regions of separate random number generation and Slepian-Wolf coding for correlated stationary memoryless sources.

Corollary 3.14 may be called *random number generation with side information at the tester*, which, to some extent, is similar to the problem of source coding with side information at the decoder [25, 26]. A similar result, which implicitly includes Corollary 3.14, was first proved in [21].

One simple application of Corollary 3.14 is the generation of random seed. When extracting randomness from a source using a seeded extractor, we sometimes have to generate a random seed from the same source. Loosely speaking, the source consists of two parts,  $X$  and  $Y$ , where  $X$  is used to generate a seed in  $\mathcal{S}$  and  $Y$  is used to generate random numbers. The possible dependence between  $X$  and  $Y$  is however annoying. Corollary 3.14 tells us that if  $n^{-1} \ln |\mathcal{S}| < \underline{\mathcal{H}}(X | Y)$  for some  $n$  large enough, then there exists a good extractor  $\varphi^{(n)}$  such that the random seed  $\varphi^{(n)}(X^{(n)})$  is not only close to uniform but also almost independent of  $Y^{(n)}$ .

The implication of Corollary 3.14 is far-reaching, and in particular, it is related to exposure-resilient cryptography [27], as the following example shows.

**Example 3.15.** Let  $X^{(n)}$  be an  $n$ -bit binary string uniformly distributed over  $\{0, 1\}^n$ . Let  $F^{(n)}$  be an exposure function that randomly exposes all but  $k_n = \lceil \alpha n \rceil$  bits of  $X^{(n)}$  with  $\alpha \in (0, 1)$ . Exactly speaking,  $F^{(n)}$  is a random map  $\{0, 1\}^n \rightarrow \{0, 1, \epsilon\}^n$  given by

$$x = x_1 x_2 \cdots x_n \mapsto F_1^{(n)}(x) F_2^{(n)}(x) \cdots F_n^{(n)}(x)$$

and

$$F_i^{(n)}(x) = \begin{cases} \epsilon & i \in I_n, \\ x_i & \text{otherwise,} \end{cases}$$

where  $I_n$  is a random set uniformly distributed over the collection of all  $k_n$ -size subsets of  $\llbracket 1, n \rrbracket$ . Let  $Y^{(n)} = F^{(n)}(X^{(n)})$ , and then  $(X, Y) = (X^{(n)}, Y^{(n)})_{n=1}^\infty$  forms a pair of correlated sources. Given the side information  $Y^{(n)}$ , it is clear that all but  $k_n$  bits of  $X^{(n)}$  are fixed. Our question is how much randomness we can extract from  $X$  given  $Y$ . By Corollary 3.14, there is a good extractor sequence  $(\varphi^{(n)} : \{0, 1\}^n \rightarrow \mathcal{Y}^{(n)})_{n=1}^\infty$  for any rate not larger than

$$\underline{\mathcal{H}}(X | Y) = \alpha \ln 2.$$

A direct usage of Lemma 3.8 further shows that

$$d(X^{(n)} | \varphi^{(n)}) \leq e^{-n(\alpha \ln 2 - R(\varphi^{(n)}) - o(1))/2}.$$

Note that  $d(X^{(n)} | \varphi^{(n)})$  is the average of  $d(\varphi^{(n)}(Z^{(n)}), U_{\mathcal{Y}^{(n)}})$  over the set of all  $n$ -bit binary strings  $Z^{(n)}$  in which  $n - k_n$  bits are fixed and the rest are uniformly distributed, where  $Z^{(n)}$  is identified with its distribution and hence the number of all  $Z^{(n)}$ 's is  $\binom{n}{k_n} 2^{n-k_n}$ .

Owing to Theorem 3.12 (as well as its special case [5, Theorem 2.2.2] for single source), the spectral inf-entropy rate becomes the most natural measure of intrinsic randomness (for fixed-rate randomness extraction), but unfortunately, it does not have a simple finite-length counterpart. A natural candidate is

$$\underline{H}_\epsilon(X) := \sup \left\{ r : P \left\{ \ln \frac{1}{P_X(X)} < r \right\} \leq \epsilon \right\}$$

(loosely speaking, the minimum of  $-\ln P_X(X)$  with confidence level  $1 - \epsilon$ ). Note that  $\underline{H}_0(X)$  is just the *min-entropy*  $\min\{-\ln P_X(X)\}$ , which not only reveals the relation between the min-entropy and the spectral inf-entropy rate but also explains why the min-entropy usually underestimates the intrinsic randomness of a source.

## 4 Refined Analysis

By Theorem 3.12 we established the complete characterization of the achievable rate region  $\mathcal{R}(X)$ , which is however not enough for practice because our proof is based on the average performance of 1-random-bin maps and hence does not give an explicit construction of extractors. In order to provide useful hints for designing extractors, we need a refined analysis for some important kinds of 1-random-bin maps, which is the task of this section.

Markov's inequality can certainly help us relate the performance of individual maps with the average performance of whole ensemble, but it is too weak. What we will use is McDiarmid's inequality [28], a concentration inequality based on the martingale approach.

The first object of study is the pure-random-bin map.

**Lemma 4.1.** *Let  $X = X_{[0,m]}$  be an  $(m+1)$ -tuple of random elements in  $\Pi(\mathcal{X})$  with  $\mathcal{X} = \mathcal{X}_{[0,m]}$ . Let  $\Phi = (\Phi_i : \mathcal{X}_i \rightarrow \mathcal{Y}_i)_{i \in [1,m]}$  be an  $m$ -tuple of pure-random-bin extractors. Then, for  $r > 1$  and  $s > 0$ ,*

$$P \left\{ d(X | \Phi) \geq P\{(T_X(X), \hat{T}_X(X)) \notin A_r(\mathcal{Y}) \times I_{rs}^m\} + \frac{\sqrt{2^m - 1} + \sqrt{2m}}{2} r^{-1/2} \right\} \leq e^{-s},$$

where  $d(\cdot | \cdot)$ ,  $T_X$ ,  $A_r$ , and  $I_t$  are defined by (4), (5), (6), and (7), respectively, and

$$\hat{T}_X(x) := (\hat{T}_X^{(i)}(x))_{i \in [1,m]} = \left( \frac{1}{P_{X_i}(x_i)} \right)_{i \in [1,m]}. \quad (8)$$

*Proof.* Let  $\lambda = P(G)$  with  $G = \{(T_X(X), \hat{T}_X(X)) \in A_r(\mathcal{Y}) \times I_{rs}^m\}$ . Since the lemma holds trivially for  $\lambda = 0$ , we suppose that  $\lambda > 0$ . Then

$$P_X = \lambda P_V + (1 - \lambda) P_W,$$

where  $P_V = P_{X|1_G=1}$  and  $P_W = P_{X|1_G=0}$ . In the same vein of the proof of Lemma 3.8, we have

$$d(X | \Phi) \leq \lambda d(V | \Phi) + 1 - \lambda, \quad (9)$$

$$P_V \leq \lambda^{-1} P_X, \quad (10)$$

$$\mathbb{E}[d(V | \Phi)] \leq \frac{\sqrt{2^m - 1}}{2\lambda} r^{-1/2}. \quad (11)$$

Since each  $\Phi_i$  is a pure-random-bin map,  $d(V \mid \Phi)$  is in fact a function of the independent random elements

$$((\Phi_1(u))_{u \in \mathcal{X}_1}, (\Phi_2(u))_{u \in \mathcal{X}_2}, \dots, (\Phi_m(u))_{u \in \mathcal{X}_m})$$

and for any  $i \in \llbracket 1, m \rrbracket$  and  $u \in \mathcal{X}_i$ ,

$$\sup_{\substack{\varphi, \psi \in \Phi(\Omega): \\ (\varphi_j = \psi_j)_{j \neq i} \\ \varphi_i|_{\{u\}^c} = \psi_i|_{\{u\}^c}}} |d(V \mid \varphi) - d(V \mid \psi)| \leq P_{V_i}(u).$$

Using McDiarmid's inequality [28, Lemma 1.2 and Theorem 6.7], we obtain

$$P\{d(V \mid \Phi) - \mathbb{E}[d(V \mid \Phi)] \geq \delta\} \leq e^{-2\delta^2/\chi}$$

with  $\chi = \sum_{i=1}^m \sum_{u \in \mathcal{X}_i} P_{V_i}(u)^2$  and  $\delta > 0$ . Choosing  $\delta = \sqrt{2s\chi}/2$ , we have

$$P\left\{d(V \mid \Phi) - \mathbb{E}[d(V \mid \Phi)] \geq \frac{\sqrt{2s\chi}}{2}\right\} \leq e^{-s}$$

and

$$\begin{aligned} s\chi &= s \sum_{i=1}^m \sum_{u \in \mathcal{X}_i} P_{V_i}(u)^2 \\ &\leq s \sum_{i=1}^m \sum_{u \in X_i(G)} \left( \sum_{(a,v) \in \mathcal{X}_{\{0\} \cup \{i\}^c}} \frac{P_{X_0 X_i X_{\{i\}^c}}(a, u, v)}{\lambda} \right)^2 \end{aligned} \quad (12)$$

$$\begin{aligned} &\leq \lambda^{-2} s \sum_{i=1}^m \sum_{u \in X_i(G)} P_{X_i}(u)^2 \\ &\leq m\lambda^{-2} r^{-1}, \end{aligned} \quad (13)$$

where (12) follows from  $P\{V_i \in X_i(G)\} = 1$  and (10), and (13) follows from a property of elements in  $X_i(G)$ , that is,  $1/P_{X_i}(u) > rs$  for  $u \in X_i(G)$ . This implies that

$$P\left\{d(X \mid \Phi) \geq \lambda \mathbb{E}[d(V \mid \Phi)] + \frac{\sqrt{2m}}{2} r^{-1/2} + 1 - \lambda\right\} \leq e^{-s},$$

which combined with (11) yields the desired result.  $\square$

An immediate consequence of Lemma 4.1 is the next theorem.

**Theorem 4.2.** *Let  $X = (X_{\llbracket 0, m \rrbracket}^{(n)})_{n=1}^\infty$  be an  $(m+1)$ -tuple of correlated sources and  $\Phi^{(n)} = \Phi_{\llbracket 1, m \rrbracket}^{(n)}$  an  $m$ -tuple of pure-random-bin extractors such that  $(\lim_{n \rightarrow \infty} \mathbb{R}(\Phi_i^{(n)}))_{i \in \llbracket 1, m \rrbracket}$  is an interior point of  $\mathcal{R}_i(X)$ . Then, for any  $\epsilon_1, \epsilon_2 > 0$ , there exists an integer  $N = N(\epsilon_1, \epsilon_2)$  such that*

$$P\{d(X^{(n)} \mid \Phi^{(n)}) \geq \epsilon_1\} \leq e^{-e^{n(\min_{1 \leq i \leq m} \mathcal{H}(X_i) - \epsilon_2)}}$$

for all  $n \geq N$ .

Loosely speaking, the fraction of bad sample maps of a pure-random-bin extractor converges to zero (as  $n$  goes to infinity) in a speed of double exponential function as long as every source to be extracted has a nonzero spectral inf-entropy rate. In other words, most maps are good extractors, since a pure-random-bin map is uniformly distributed over the set of all maps with the same domain and codomain.

**Example 4.3** (cf. [27, 29]). *Let us resume the discussion of Example 3.15, with also the same definitions and notations. Lemma 4.1 shows that, for  $\nu > 0$ , the pure-random-bin extractor  $\Phi^{(n)} : \{0, 1\}^n \rightarrow \mathcal{Y}^{(n)}$  satisfies*

$$P\{d(\Phi^{(n)}(Z^{(n)}), U_{\mathcal{Y}^{(n)}}) \geq 1.21e^{-n(\alpha \ln 2 - \rho_n - \nu)/2}\} \leq e^{-e^{n\rho_n}}.$$

where  $\rho_n = \max\{R(\Phi^{(n)}), \ln(2n)/n\}$ . Since the number of all  $Z^{(n)}$ 's is  $\binom{n}{k_n} 2^{n-k_n}$ , we have

$$\binom{n}{k_n} 2^{n-k_n} e^{-e^{n\rho_n}} < 2^{2n} e^{-2n} = o(1),$$

so that there is an extractor  $\varphi^{(n)} : \mathcal{X}^n \rightarrow \mathcal{Y}^{(n)}$  satisfying

$$d(\varphi^{(n)}(Z^{(n)}), U_{\mathcal{Y}^{(n)}}) = e^{-n(\alpha \ln 2 - \rho_n - o(1))/2}$$

for all  $Z^{(n)}$ 's. This pointwise performance is clearly much stronger than the average performance in Example 3.15. If using Lemma 4.4 instead of Lemma 4.1, we can further conclude that there is an equal-bin extractor achieving this pointwise performance.

The next object of study is the equal-random-bin map, which has a similar concentration phenomenon to the pure-random-bin map.

**Lemma 4.4.** *Let  $X = X_{[0,m]}$  be an  $(m+1)$ -tuple of random elements in  $\Pi(\mathcal{X})$  with  $\mathcal{X} = \mathcal{X}_{[0,m]}$ . Let  $\Phi = (\Phi_i : \mathcal{X}_i \rightarrow \mathcal{Y}_i)_{i \in [1,m]}$  be an  $m$ -tuple of equal-random-bin extractors. Then, for  $r > 1$  and  $s > 0$ ,*

$$P \left\{ d(X \mid \Phi) \geq P\{(T_X(X), \hat{T}_X(X)) \notin A_r(\mathcal{Y}) \times I_{rs}^m\} + \frac{\sqrt{2^m - 1} + 2\sqrt{2m}}{2} r^{-1/2} \right\} \leq e^{-s},$$

where  $d(\cdot \mid \cdot)$ ,  $T_X$ ,  $\hat{T}_X$ ,  $A_r$ , and  $I_t$  are defined by (4), (5), (8), (6), and (7), respectively. If  $m = 1$  and  $X_1$  is independent of  $X_0$ , then we have a slightly improved result:

$$P \left\{ d(X \mid \Phi) \geq P\{(T_X(X), \hat{T}_X(X)) \notin A_r(\mathcal{Y}) \times I_{rs}^m\} + \frac{1 + \sqrt{2}}{2} e^{-n\gamma/2} \right\} \leq e^{-s}.$$

*Proof.* Let  $\lambda = P(G)$  with  $G = \{(T_X(X), \hat{T}_X(X)) \in A_r(\mathcal{Y}) \times I_{rs}^m\}$ . In the same vein of the proof of Lemma 4.1, we have (9)–(11).

Since each  $\Phi_i$  is an equal-random-bin map, it can be expressed as

$$\Phi_i = U_{\mathcal{S}_{\mathcal{Y}_i}} \circ \varphi_i \circ U_{\mathcal{S}_{\mathcal{X}_i}}$$

where  $\varphi_i$  is an arbitrary fixed equal-bin map from  $\mathcal{X}_i$  onto  $\mathcal{Y}_i$ . Since the probability distribution of  $U_{\mathcal{Y}_i}$  is invariant under  $U_{\mathcal{S}_{\mathcal{Y}_i}}$ , we further assume that  $\Phi_i = \varphi_i \circ U_{\mathcal{S}_{\mathcal{X}_i}}$ . Then  $d(V \mid \Phi)$  becomes

a function of the random element  $((U_{S_{\mathcal{X}_i}}(u))_{u \in \mathcal{X}_i})_{i \in \llbracket 1, m \rrbracket}$ . Let  $t = (t_i : \llbracket 1, |\mathcal{X}_i| \rrbracket \rightarrow \mathcal{X}_i)_{i \in \llbracket 1, m \rrbracket}$  be an  $m$ -tuple of one-to-one maps such that  $P_{V_i}(t_i(\ell))$  is nonincreasing in  $\ell$  for all  $i \in \llbracket 1, m \rrbracket$ , so that  $((U_{S_{\mathcal{X}_i}}(u))_{u \in \mathcal{X}_i})_{i \in \llbracket 1, m \rrbracket}$  can be ordered as the following sequence:

$$(U_{S_{\mathcal{X}_1}}(t_1(1)), \dots, U_{S_{\mathcal{X}_1}}(t_1(|\mathcal{X}_1|)), \dots, U_{S_{\mathcal{X}_m}}(t_m(1)), \dots, U_{S_{\mathcal{X}_m}}(t_m(|\mathcal{X}_m|))).$$

Let  $\mathcal{Z} = \bigcup_{i=1}^m \{(i, j) : j \in \llbracket 1, |\mathcal{X}_i| \rrbracket\}$  and  $\mathcal{S} = \prod_{i=1}^m S_{\mathcal{X}_i}$ . For  $\pi, \pi' \in \mathcal{S}$ , we say  $\pi =_{k, \ell} \pi'$  if  $\pi_i = \pi'_i$  for all  $i < k$  and  $\pi_k(t_k(j)) = \pi'_k(t_k(j))$  for all  $j \leq \ell$ . For  $(k, \ell) \in \mathcal{Z}$  and  $\pi \in \mathcal{S}$ , we define

$$f_{k, \ell}(\pi) = \mathbb{E}[d(V \mid \Phi) \mid (U_{S_{\mathcal{X}_i}})_{i \in \llbracket 1, m \rrbracket} =_{k, \ell} \pi].$$

From Proposition A.3 it follows that

$$|f_{k, \ell}(\pi) - f_{k, \ell}(\pi')| \leq 2P_{V_k}(t_k(\ell)) \quad (14)$$

for all  $\pi, \pi' \in \mathcal{S}$  with  $\pi =_{k, \ell-1} \pi'$ . Using Proposition A.4, we obtain

$$P\{d(V \mid \Phi) - \mathbb{E}[d(V \mid \Phi)] \geq \delta\} \leq e^{-\delta^2/(2\chi)}$$

with  $\chi = \sum_{i=1}^m \sum_{u \in \mathcal{X}_i} P_{V_i}(u)^2$  and  $\delta > 0$ . Choosing  $\delta = \sqrt{2s\chi}$ , we have

$$P\left\{d(V \mid \Phi) - \mathbb{E}[d(V \mid \Phi)] \geq \sqrt{2s\chi}\right\} \leq e^{-s}$$

and (13) in the same vein of the proof of Lemma 4.1. This implies that

$$P\left\{d(X \mid \Phi) \geq \lambda \mathbb{E}[d(V \mid \Phi)] + \sqrt{2mr}^{-1/2} + 1 - \lambda\right\} \leq e^{-s},$$

which together with (11) establishes the main part of the lemma. The last statement of the lemma can simply be proved by replacing (14) with

$$|f_{1, \ell}(\pi) - f_{1, \ell}(\pi')| \leq P_{V_1}(t_1(\ell))$$

according to the improved bound of Proposition A.3. □

An immediate consequence of Lemma 4.4 is the next theorem.

**Theorem 4.5.** *Let  $X = (X_{\llbracket 0, m \rrbracket}^{(n)})_{n=1}^\infty$  be an  $(m+1)$ -tuple of correlated sources and  $\Phi^{(n)} = \Phi_{\llbracket 1, m \rrbracket}^{(n)}$  an  $m$ -tuple of equal-random-bin extractors such that  $(\lim_{n \rightarrow \infty} \mathbf{R}(\Phi_i^{(n)}))_{i \in \llbracket 1, m \rrbracket}$  is an interior point of  $\mathcal{R}_1(X)$ . Then, for any  $\epsilon_1, \epsilon_2 > 0$ , there exists an integer  $N = N(\epsilon_1, \epsilon_2)$  such that*

$$P\{d(X^{(n)} \mid \Phi^{(n)}) \geq \epsilon_1\} \leq e^{-e^{n(\min_{1 \leq i \leq m} \mathbf{H}(X_i) - \epsilon_2)}}$$

for all  $n \geq N$ .



## 5 An Open and Transparent Lottery Scheme for Resource Allocation

In today's economic systems, goods (and services) are usually allocated by a market. In most cases market works well, but it fails when there is a shortage of a good (which is a “public good” in many cases and) whose allocation is related to social justice issues. Examples of such a good include traffic, fresh air, affordable housing, education, and so on. One feasible solution to the allocation of these resources is lottery, operated by an authority, such as a local government. A great advantage of lottery is its simple fairness based on randomness, so it is widely used in many countries, e.g., affordable housing lottery in New York [30], car license-plate lottery in Beijing [31].

However, can we trust in the randomness of a lottery? In particular, can we ensure that there is no fraud in the process of a lottery? Now we will present an open and transparent lottery scheme that is immune to any kinds of fraud under some necessary conditions.

Suppose a public source  $X$  and a lottery of  $k$  participants with an  $l$ -tuple  $s = (s_i)_{i \in [1, \ell]}$  of shares  $s_i$  of the total prize (which can be a fixed amount of cash or goods), where  $1 \leq \ell \leq k$ .  $\ell$  of  $k$  participants are to be chosen as the winners of prizes  $s_1, s_2, \dots, s_\ell$ , respectively, in a random order by the lottery based on the intrinsic randomness of  $X$ .

To secure the process of lottery, we require that  $X$  be not in the control of anyone, or at least, that any valid fraud on  $X$  cost the cheater much more than the prize of the lottery. A natural phenomenon that can be easily witnessed by anyone is one kind of candidates for  $X$ . Another kind of candidates is a man-made public event involving interactions among a large number of people. Clearly, no one can easily control such an event. Examples of the above two kinds of sources are weather and share prices, respectively.

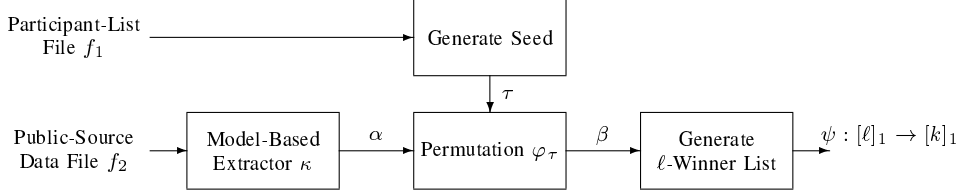
The details of our lottery scheme is as follows:

1. Generate an electronic file  $f_1$  containing an ordered list of all participants (numbered from 1 to  $k$ ) as well as all necessary information for identifying the participants and verifying their eligibility. Compute the hash value of  $f_1$  using a cryptographic hash function (e.g., SHA-1) and disclose this value as well as the number  $k$  to the public.
2. Choose and announce in public a future time interval  $[t_1, t_2]$  of  $X$  for lottery. An estimation is needed to ensure that source  $X$  outputs enough randomness for lottery during the chosen period.
3. Record the data of  $X$  from  $t_1$  to  $t_2$  into an electronic file  $f_2$ .
4. Compute the lottery result by a software extracting randomness from  $f_2$  with also a seed generated from  $f_1$ . Declare the list of winners and disclose to the public the files  $f_1$  and  $f_2$  as well as the source code of the software, so that everyone can verify the lottery.

Since the statistics of  $X$  may be very complicated, the success of our scheme depends on the performance of the extractor used by the software. Corollary 3.14 and Lemma 3.8 tell us that, whatever  $X$  is, there is an “almost-blind” good extractor for any rate  $R$  below the information-theoretic bound  $\underline{H}(X | Y)$  of  $X$  given any side information  $Y$ .<sup>4</sup> In particular, the output of extractor is almost independent of  $Y$ . Since  $X$  can be regarded as the output of a virtual channel with input

---

<sup>4</sup> Lemma 3.8 shows that a 1-random-bin extractor works well at any rate  $R < \underline{H}(X | Y)$ , regardless of the actual probability distribution of the correlated sources  $(X, Y)$ . Such an extractor ensemble is said to be “blind” or “universal”, but there is no individual extractor having such a property. In fact, for any extractor  $\varphi^{(n)} : \mathcal{X}^n \rightarrow$



**Figure 2:** The proposed framework of lottery algorithm.

$Y$ , the quantity  $\underline{H}(X | Y)$  actually shows the ability of the potential adversary associated with  $Y$ . This then implies that our lottery scheme is immune to any attack based on any side information  $Y$  with  $\underline{H}(X | Y) > R$ .

Having introduced the whole lottery scheme, we proceed to design the lottery algorithm. We will only present a basic framework, leaving the details to the author’s future work.

The goal of this algorithm is to extract randomness from provided data and then to generate a uniformly distributed one-to-one map from  $\llbracket 1, \ell \rrbracket$  into  $\llbracket 1, k \rrbracket$ , i.e., the  $\ell$ -winner list of the lottery. Fig. 2 shows the basic idea of the algorithm, which consists of three steps:

1. Process the data file  $f_2$  by a model-based extractor  $\kappa$ , which is a traditional compression algorithm but with necessary modifications for random number generation, i.e., discarding any output that is possibly structured. Examples of this idea are [6] and [32]. The output of this step is denoted by  $\alpha$ , a binary string.
2. Permute  $\alpha$  by a permutation  $\varphi_\tau$  of  $\{0, 1\}^{|\alpha|}$ , which is chosen from some collection of permutations according to the seed  $\tau$  generated from  $f_1$  (by a hash function different from the hash function already used in the first step of the lottery scheme). The output of this step is denoted by  $\beta$ , also a binary string. If well designed, any truncation of  $\varphi_\tau$  can be a good extractor of that rate, or equivalently,  $\varphi_\tau$  is a good rateless extractor whose existence is confirmed by the theoretical performance of equal-random-bin maps (Theorem 4.5).
3. Generate from  $\beta$  a one-to-one map  $\psi : \llbracket 1, \ell \rrbracket \rightarrow \llbracket 1, k \rrbracket$ .

We may, for example, use the Fisher-Yates shuffle algorithm [33, Algorithm P at p. 145] and the interval algorithm [3] to generate from  $\beta$  a random permutation  $\pi$  of  $\llbracket 1, k \rrbracket$ . The restriction of  $\pi$  to  $\llbracket 1, \ell \rrbracket$  then gives a one-to-one map  $\psi$  from  $\llbracket 1, \ell \rrbracket$  into  $\llbracket 1, k \rrbracket$ . The actual data consumed for generating  $\pi$  are denoted by  $\beta'$ , a truncation of  $\beta$ . Since  $\varphi_\tau$  is a good rateless extractor,  $\beta'$  is surely close to uniform (provided that data  $f_2$  has enough intrinsic randomness). The length of  $\beta'$  is about  $\log_2(k!) \approx (k + \frac{1}{2}) \log_2 k - k \log_2 e$  bits, which combined with the spectral inf-entropy rate of  $X$  can be used to determine the time interval  $[t_1, t_2]$  for lottery.

By a careful arrangement of the algorithm, we can ensure that  $\psi$ , i.e.,  $\pi|_{\llbracket 1, \ell \rrbracket}$  for any  $\ell \in \llbracket 1, k \rrbracket$ , always depends on the whole data  $\beta'$ , so that the actual probability distribution of  $\psi$  will not deviate too much even if there are severe local defects in  $\beta'$ .<sup>5</sup>

Let us close this section with a summary of our lottery scheme. First and foremost, it is open and transparent, and to some extent, immune to all kinds of fraud. Second, it can largely reduce

---

$\{0, 1\}$ , the general source  $X = (U_{(\varphi^{(n)})^{-1}(y_n)})_{n=1}^\infty$  with  $y_n$  maximizing  $|(\varphi^{(n)})^{-1}(y_n)|$  satisfies  $\underline{H}(X) = \ln |\mathcal{X}|$  and  $d(\varphi^{(n)}(X^{(n)}), U_{\{0,1\}}) = \frac{1}{2}$ . For this reason, we only claim the existence of an “almost-blind” good extractor, which is confirmed by Theorems 4.2 and 4.5.

<sup>5</sup> There’s no such thing as a free lunch, because we spend more randomness on  $\psi$  than needed. In other words, this algorithm has an intrinsic extractor (cf. Theorem 4.5).

the cost of lottery, such as the cost for notarization, which is sometimes a necessary part of a lottery. Its cost-saving feature will become more prominent if it is further standardized, e.g., using a standard open-source software as well as standard public data sources for lottery. Third, the proposed framework of lottery algorithm is in fact a combination of knowledge-based approaches (Model-based extractor  $\kappa$ ) and non-knowledge-based approaches (Permutation  $\varphi_\tau$ ). This idea is so fundamental that it is useful for any design of random number generator. Using a permutation as a rateless extractor is also an interesting idea. Finding such a permutation then becomes an important research issue. Possible candidates include all existing encryption algorithms.

## 6 Conclusion

In this paper, we investigated the problem of separate random number generation from correlated sources with side information at the tester. Tight one-shot lower and upper performance bounds, as well as the achievable rate region, are obtained. By a refined analysis, we further proved the doubly-exponential concentration phenomenon of the performance of two kinds of random extractors: the pure-random-bin extractor and the equal-random-bin extractor. As an application of these results, we presented an open and transparent lottery scheme for resource allocation, which may be the first lottery scheme immune to all kinds of fraud.

The generation of random numbers is a fundamental issue in information science, extensively related to our understanding of randomness, knowledge, and intelligence. This paper, in the author's view, discovered and proved some very fundamental facts that are necessary for the theory of universal random number generation, which may finally be established via the approaches of information theory, coding theory, decision theory, and computational complexity theory.

## A Useful Facts

**Proposition A.1.** *If  $V$ ,  $V'$ ,  $W$ , and  $W'$  are random elements in the measurable space  $(\mathcal{X}, \mathfrak{X})$ , then*

$$d(\lambda P_V + (1 - \lambda)P_W, \lambda P_{V'} + (1 - \lambda)P_{W'}) \leq \lambda d(P_V, P_{V'}) + (1 - \lambda) d(P_W, P_{W'})$$

for all  $\lambda \in [0, 1]$ .

*Proof.*

$$\begin{aligned} & d(\lambda P_V + (1 - \lambda)P_W, \lambda P_{V'} + (1 - \lambda)P_{W'}) \\ &= \sup_{A \in \mathfrak{X}} |(\lambda P_V + (1 - \lambda)P_W)(A) - (\lambda P_{V'} + (1 - \lambda)P_{W'})(A)| \\ &= \sup_{A \in \mathfrak{X}} |\lambda(P_V(A) - P_{V'}(A)) + (1 - \lambda)(P_W(A) - P_{W'}(A))| \\ &\leq \sup_{A \in \mathfrak{X}} [\lambda|P_V(A) - P_{V'}(A)| + (1 - \lambda)|P_W(A) - P_{W'}(A)|] \\ &\leq \lambda d(P_V, P_{V'}) + (1 - \lambda) d(P_W, P_{W'}). \end{aligned}$$

□

**Proposition A.2.** *Let  $V$  and  $V'$  be two random elements in the measurable space  $(\mathcal{X}, \mathfrak{X})$ . Let  $f$  be a measurable mapping from  $(\mathcal{X}, \mathfrak{X})$  to  $(\mathcal{Y}, \mathfrak{Y})$ . Then  $d(P_V, P_{V'}) \geq d(P_{f(V)}, P_{f(V')})$ .*

*Proof.*

$$\begin{aligned}
d(P_{f(V)}, P_{f(V')}) &= \sup_{B \in \mathfrak{Y}} |P_{f(V)}(B) - P_{f(V')}(B)| \\
&= \sup_{B \in \mathfrak{Y}} |P_V(f^{-1}(B)) - P_{V'}(f^{-1}(B))| \\
&\leq \sup_{A \in \mathfrak{X}} |P_V(A) - P_{V'}(A)| = d(P_V, P_{V'}).
\end{aligned}$$

□

**Proposition A.3.** *Let  $\mathcal{X} = \mathcal{X}_{[0,m]}$  and  $\mathcal{Y} = \mathcal{Y}_{[1,m]}$  be two families of alphabets with  $\mathcal{X}_0$  at most countable and all the other alphabets finite. Let  $V = V_{[0,m]}$  be a random element in  $\Pi(\mathcal{X})$  and  $\Phi = (\varphi_i \circ \text{U}_{S_{\mathcal{X}_i}})_{i \in [1,m]}$  with  $\varphi_i$  an arbitrary equal-bin map from  $\mathcal{X}_i$  onto  $\mathcal{Y}_i$ .*

*Let  $t = (t_i : [1, |\mathcal{X}_i|] \rightarrow \mathcal{X}_i)_{i \in [1,m]}$  be an  $m$ -tuple of one-to-one maps such that  $P_{V_i}(t_i(\ell))$  is nonincreasing in  $\ell$  for all  $i \in [1, m]$ . Let  $\mathcal{Z} = \bigcup_{i=1}^m \{(i, j) : j \in [1, |\mathcal{X}_i|]\}$  and  $\mathcal{S} = \prod_{i=1}^m S_{\mathcal{X}_i}$ . For  $\pi, \pi' \in \mathcal{S}$ , we say  $\pi =_{k,\ell} \pi'$  if  $\pi_i = \pi'_i$  for all  $i < k$  and  $\pi_k(t_k(j)) = \pi'_k(t_k(j))$  for all  $j \leq \ell$ .*

*For  $(k, \ell) \in \mathcal{Z}$  and  $\pi \in \mathcal{S}$ , define*

$$f_{k,\ell}(\pi) = \mathbb{E}[d(V \mid \Phi) \mid (\text{U}_{S_{\mathcal{X}_i}})_{i \in [1,m]} =_{k,\ell} \pi].$$

*Then*

$$|f_{k,\ell}(\pi) - f_{k,\ell}(\pi')| \leq 2P_{V_k}(t_k(\ell))$$

*for all  $\pi, \pi' \in \mathcal{S}$  with  $\pi =_{k,\ell-1} \pi'$ . In particular, if  $P_{V_k}$  has a nonincreasing order  $t_k$  independent of  $V_{\{0\} \cup \{k\}^c}$ , that is,  $P_{V_0 V_k V_{\{k\}^c}}(a, t_k(\ell), v)$  is nonincreasing in  $\ell$  for all  $a \in \mathcal{X}_0$  and  $v \in \mathcal{X}_{\{k\}^c}$ , then the coefficient 2 of this upper bound can be replaced with 1.*

*Proof.* Let  $\mathcal{S}' = \{\sigma \in \mathcal{S} : \sigma =_{k,\ell} \pi\}$ . Then

$$\begin{aligned}
|f_{k,\ell}(\pi) - f_{k,\ell}(\pi')| &= \left| \sum_{\substack{\sigma \in \mathcal{S}' \\ \sigma' = \iota \circ \sigma}} \left( P_{\text{U}_{\mathcal{S}'}}(\sigma) d(P_{V_0((\varphi_i \circ \sigma_i)(V_i))_{i \in [1,m]}}, P_{V_0} P_{\text{U}_{\Pi(\mathcal{Y})}}) \right. \right. \\
&\quad \left. \left. - P_{\text{U}_{\mathcal{S}'}}(\sigma') d(P_{V_0((\varphi_i \circ \sigma'_i)(V_i))_{i \in [1,m]}}, P_{V_0} P_{\text{U}_{\Pi(\mathcal{Y})}}) \right) \right| \\
&= \left| \sum_{\substack{\sigma \in \mathcal{S}' \\ \sigma' = \iota \circ \sigma}} P_{\text{U}_{\mathcal{S}'}}(\sigma) \left( d(P_{V_0((\varphi_i \circ \sigma_i)(V_i))_{i \in [1,m]}}, P_{V_0} P_{\text{U}_{\Pi(\mathcal{Y})}}) \right. \right. \\
&\quad \left. \left. - d(P_{V_0((\varphi_i \circ \sigma'_i)(V_i))_{i \in [1,m]}}, P_{V_0} P_{\text{U}_{\Pi(\mathcal{Y})}}) \right) \right| \\
&\leq \sum_{\substack{\sigma \in \mathcal{S}' \\ \sigma' = \iota \circ \sigma}} P_{\text{U}_{\mathcal{S}'}}(\sigma) \left| d(P_{V_0((\varphi_i \circ \sigma_i)(V_i))_{i \in [1,m]}}, P_{V_0} P_{\text{U}_{\Pi(\mathcal{Y})}}) \right. \\
&\quad \left. - d(P_{V_0((\varphi_i \circ \sigma'_i)(V_i))_{i \in [1,m]}}, P_{V_0} P_{\text{U}_{\Pi(\mathcal{Y})}}) \right| \\
&\leq \sum_{\substack{\sigma \in \mathcal{S}' \\ \sigma' = \iota \circ \sigma}} P_{\text{U}_{\mathcal{S}'}}(\sigma) \left| d(P_{V_0((\varphi_i \circ \sigma_i)(V_i))_{i \in [1,m]}}, P_{V_0((\varphi_i \circ \sigma'_i)(V_i))_{i \in [1,m]}}) \right|
\end{aligned}$$

$$\leq \sum_{\substack{\sigma \in \mathcal{S}' \\ \sigma' = \iota \circ \sigma}} P_{U_{\mathcal{S}'}}(\sigma) \left| d(P_{V_0 \sigma_k(V_k) V_{\{k\}^c}}, P_{V_0 \sigma'_k(V_k) V_{\{k\}^c}}) \right| \quad (15)$$

$$\leq P_{V_k}(t_k(\ell)) + P_{V_k}(\pi_k^{-1}(\pi'_k(t_k(\ell)))) \quad (16)$$

$$\leq 2P_{V_k}(t_k(\ell)), \quad (17)$$

where

$$\iota = (\text{id}_{\mathcal{X}_1}, \dots, \text{id}_{\mathcal{X}_{k-1}}, (\pi_k(t_k(\ell)) \pi'_k(t_k(\ell))), \text{id}_{\mathcal{X}_{k+1}}, \dots, \text{id}_{\mathcal{X}_m}),$$

(15) follows from Proposition A.2, (16) is a straightforward computation by definition, and (17) follows from the nonincreasing-probability property of  $t_k$ . The proof is complete by noting that if  $P_{V_k}$  has a nonincreasing order  $t_k$  independent of  $V_{\{0\} \cup \{k\}^c}$ , then step (16) can be replaced with

$$P_{V_k}(t_k(\ell)) - P_{V_k}(\pi_k^{-1}(\pi'_k(t_k(\ell)))),$$

so that the last upper bound becomes  $P_{V_k}(t_k(\ell))$ .  $\square$

**Proposition A.4** (A modified version of [28, Corollary 6.10]). *Let  $Z = (Z_i)_{i=1}^n$  be an  $n$ -tuple of random variables in  $\Pi(A)$  with  $A = (A_i)_{i=1}^n$ . Let  $D_i$  be the corresponding subset of  $\Pi(A_{[1,i]})$  such that  $P_{Z_{[1,i]}}(D_i^c) = 0$ . Let  $f : \Pi(A) \rightarrow \mathbf{R}$  be a measurable function such that  $f(Z)$  is integrable. Suppose that there are constants  $c_1, \dots, c_n$  so that*

$$|\mathbf{E}[f(Z) \mid Z_{[1,k]} = z_{[1,k]}] - \mathbf{E}[f(Z) \mid Z_{[1,k]} = z'_{[1,k]}]| \leq c_k$$

for each  $k \in [1, n]$  and any  $z_{[1,k]}, z'_{[1,k]} \in D_k$  satisfying  $z_{[1,k-1]} = z'_{[1,k-1]}$ . Then for any  $t > 0$ ,

$$P\{|f(Z) - \mathbf{E} f(Z)| \geq t\} \leq 2e^{-2t^2 / \sum_{i=1}^n c_i^2}.$$

*Proof.* See [28, Corollary 6.10].  $\square$

## B Omitted Stuff

*Proof of Theorem 3.12.* 1) Direct part: Suppose that  $R \in \mathcal{R}_i(X)$ . For  $\gamma \in (0, \frac{1}{2} \min_{i: R_i > 0} R_i)$ , we define

$$\mathcal{Y}_i^{(n)} = [1, e^{n \max\{R_i - 2\gamma, 0\}}]$$

and use 1-random-bin maps (see Examples 3.2–3.5) as the  $m$ -tuple  $\Phi^{(n)}$  of random extractors. Lemma 3.8 with  $r = e^{n\gamma}$  shows that

$$\mathbf{E}[d(X^{(n)} \mid \Phi^{(n)})] \leq P\{T_{X^{(n)}}(X^{(n)}) \notin A_r(\mathcal{Y}^{(n)})\} + \frac{\sqrt{2^m - 1}}{2} e^{-n\gamma/2},$$

so that

$$\lim_{n \rightarrow \infty} \mathbf{E}[d(X^{(n)} \mid \Phi^{(n)})] \leq \lim_{n \rightarrow \infty} P\{T_{X^{(n)}}(X^{(n)}) \notin A_r(\mathcal{Y}^{(n)})\} = 0$$

by the definitions of  $\mathcal{R}_i(X)$  and  $\underline{\mathcal{H}}(X_S \mid X_0)$ . This implies that there is an  $m$ -tuple  $\varphi^{(n)}$  of extractors such that  $R(\varphi_i^{(n)}) = n^{-1} \ln \lfloor e^{n \max\{R_i - 2\gamma, 0\}} \rfloor$  for all  $i \in [1, m]$  and  $\lim_{n \rightarrow \infty} d(X^{(n)} \mid \varphi^{(n)}) = 0$ .

In order to complete the proof, we use the diagonal line argument. Choose a decreasing sequence  $(\gamma_k)_{k=1}^\infty$  converging to zero. For each  $\gamma_k$ , we repeat the above argument to obtain an  $m$ -tuple of

extractors, denoted  $\varphi^{(n,k)}$ , and we denote by  $N(k)$  the least integer such that  $d(X^{(n)} | \varphi^{(n,k)}) \leq \gamma_k$  for all  $n \geq N(k)$ . Define the new  $m$ -tuple  $\psi^{(n)}$  of extractors by

$$\psi^{(n)} = \varphi^{(n, \max(\{1\} \cup \{k: N(k) \leq n\}))}.$$

Then

$$\liminf_{n \rightarrow \infty} R(\psi_i^{(n)}) \geq \sup_k \lim_{n \rightarrow \infty} R(\varphi_i^{(n,k)}) = R_i \quad \text{for all } i \in \llbracket 1, m \rrbracket$$

and

$$\limsup_{n \rightarrow \infty} d(X^{(n)} | \psi^{(n)}) \leq \inf_k \gamma_k = 0.$$

Therefore,  $R \in \mathcal{R}(X)$ .

2) Converse part: Suppose that  $R \in \mathcal{R}(X)$ . Then there exists an  $m$ -tuple  $\varphi^{(n)}$  of extractors such that  $\liminf_{n \rightarrow \infty} R(\varphi_i^{(n)}) \geq R_i$  for all  $i \in \llbracket 1, m \rrbracket$  and  $\lim_{n \rightarrow \infty} d(X^{(n)} | \varphi^{(n)}) = 0$ . By Lemma 3.10 with  $r = e^{-n\gamma}$ ,

$$\liminf_{n \rightarrow \infty} P\{T_{X^{(n)}}(X^{(n)}) \in A_r(\mathcal{Y}^{(n)})\} \geq \liminf_{n \rightarrow \infty} [1 - (2^m - 1)e^{-n\gamma} - d(X^{(n)} | \varphi^{(n)})] = 1$$

for any  $\gamma > 0$ . This implies that, for every nonempty  $S \subseteq \llbracket 1, m \rrbracket$  and any  $\gamma > 0$ ,

$$\begin{aligned} \Sigma(R_S) &\leq \sum_{i \in S} \liminf_{n \rightarrow \infty} R(\varphi_i^{(n)}) \leq \liminf_{n \rightarrow \infty} \sum_{i \in S} R(\varphi_i^{(n)}) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \ln |\Pi(\mathcal{Y}_S^{(n)})| < \underline{H}(X_S | X_0) + \gamma, \end{aligned}$$

so that  $\Sigma(R_S) \leq \underline{H}(X_S | X_0)$ , and therefore  $R \in \mathcal{R}_i(X)$ .  $\square$

## References

- [1] J. von Neumann, “Various techniques used in connection with random digits,” in *Monte Carlo Method, National Bureau of Standards, Applied Math. Series*, 1951, vol. 12, pp. 36–38. 1
- [2] S. Vembu and S. Verdu, “Generating random bits from an arbitrary source: Fundamental limits,” *IEEE Transactions on Information Theory*, vol. 41, no. 5, pp. 1322–1332, Sep. 1995. 1
- [3] T. S. Han and M. Hoshi, “Interval algorithm for random number generation,” *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 599–611, Mar. 1997. 1, 17
- [4] K. Visweswariah, S. R. Kulkarni, and S. Verdú, “Source codes as random number generators,” *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 462–471, Mar. 1998. 1
- [5] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer, 2003. 1, 3, 10, 12
- [6] H. Zhou and J. Bruck, “Efficient generation of random bits from finite state markov chains,” *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2490–2506, Apr. 2012. 1, 17
- [7] G. Seroussi and M. J. Weinberger, “Optimal algorithms for universal random number generation from finite memory sources,” *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1277–1297, Mar. 2015. 1

- [8] N. Nisan and A. Ta-Shma, “Extracting randomness: A survey and new constructions,” *Journal of Computer and System Sciences*, vol. 58, no. 1, pp. 148–173, Feb. 1999. 2
- [9] V. Guruswami, C. Umans, and S. Vadhan, “Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes,” *Journal of the ACM*, vol. 56, no. 4, pp. 20:1–20:34, Jun. 2009. 2
- [10] J. L. Carter and M. N. Wegman, “Universal classes of hash functions,” *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 143–154, Apr. 1979. 2, 8
- [11] T. M. Cover, “A proof of the data compression theorem of Slepian and Wolf for ergodic sources,” *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 226–228, Mar. 1975. 2
- [12] J. Radhakrishnan and A. Ta-Shma, “Bounds for dispersers, extractors, and depth-two superconcentrators,” *SIAM Journal on Discrete Mathematics*, vol. 13, no. 1, pp. 2–24, Jan. 2000. 2
- [13] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973. 2, 10
- [14] P. Delsarte, “Bilinear forms over a finite field, with applications to coding theory,” *Journal of Combinatorial Theory, Series A*, vol. 25, no. 3, pp. 226–241, 1978. 5
- [15] E. M. Gabidulin, “Theory of codes with maximum rank distance,” *Problems of Information Transmission*, vol. 21, no. 1, pp. 1–12, 1985. 5
- [16] S. Yang and T. Honold, “Good random matrices over finite fields,” *Advances in Mathematics of Communications*, vol. 6, no. 2, pp. 203–227, May 2012. 5
- [17] Y. Oohama, “Intrinsic randomness problem in the framework of Slepian-Wolf separate coding system,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E90-A, no. 7, pp. 1406–1417, Jul. 2007. 8, 9, 10
- [18] M. H. Yassaee, M. R. Aref, and A. Gohari, “Achievability proof via output statistics of random binning,” *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6760–6786, Nov. 2014. 8, 10
- [19] Y. Mansour, N. Nisan, and P. Tiwari, “The computational complexity of universal hashing,” *Theoretical Computer Science*, vol. 107, no. 1, pp. 121–133, Jan. 1993. 8, 9
- [20] M. H. Yassaee, M. R. Aref, and A. Gohari, “Non-asymptotic output statistics of random binning and its applications,” *arXiv:1303.0695 [cs, math]*, Mar. 2013, arXiv: 1303.0695. 9
- [21] M. Bloch, “Channel intrinsic randomness,” in *Proc. 2010 IEEE International Symposium on Information Theory*, Jun. 2010, pp. 2607–2611. 10, 11
- [22] T. Han and S. Verdu, “Approximation theory of output statistics,” *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993. 10
- [23] Y. Oohama, “Multiterminal random number generation in of separate coding systems,” in *Proc. IEEE Int. Symp. Information Theory*. IEEE, 2005, pp. 478–481. 10
- [24] S. Miyake and F. Kanaya, “Coding theorems on correlated general sources,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E78-A, no. 9, pp. 1063–1070, Sep. 1995. 10
- [25] A. D. Wyner, “On source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 21, no. 3, pp. 294–300, May 1975. 11



- [26] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, “Nonasymptotic and second-order achievability bounds for coding with side-information,” *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1574–1605, Apr. 2015. 11
- [27] J. Kamp and D. Zuckerman, “Deterministic extractors for bit-fixing sources and exposure-resilient cryptography,” *SIAM Journal on Computing*, vol. 36, no. 5, pp. 1231–1247, Dec. 2006. 11, 14
- [28] C. McDiarmid, “On the method of bounded differences,” in *Surveys in Combinatorics, London Mathematical Society Lecture Note Series*, 1989, vol. 141, pp. 148–188. 12, 13, 20
- [29] C.-J. Lee, C.-J. Lu, and S.-C. Tsai, “Deterministic extractors for independent-symbol sources,” *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6501–6512, Dec. 2010. 14
- [30] (2013, Aug.) HPD commissioner Wambua, HDC president Jahr and council speaker Quinn announce official launch of city’s new affordable housing lottery website. New York City Department of Housing Preservation and Development. [Online]. Available: <http://www.nyc.gov/html/hpd/html/pr2013/pr-08-20-13.shtml> 16
- [31] (2011, Jan.) Beijing starts car plate lottery. China Central Television. [Online]. Available: <http://english.cntv.cn/program/newsupdate/20110126/105041.shtml> 16
- [32] G. Seroussi and M. J. Weinberger, “Twice-universal fixed to variable-length random number generators for finite memory sources,” in *Proc. 2013 IEEE International Symposium on Information Theory*, Jul. 2013, pp. 634–638. 17
- [33] D. E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 3rd ed. Reading, Massachusetts, USA: Addison-Wesley, 1998. 17